



UNIVERSITY
of
GLASGOW

Urban, J. and Jose, J.M. (2006) Can a workspace help to overcome the query formulation problem in image retrieval? In, *28th European Conference on Information Retrieval (ECIR'06), 10-12 April 2006* Lecture Notes in Computer Science Vol 3936, pages pp. 385-396, London, UK.

<http://eprints.gla.ac.uk/3582/>

Can a Workspace Help to Overcome the Query Formulation Problem in Image Retrieval?

Jana Urban and Joemon M. Jose

Department of Computing Science, University of Glasgow,
Glasgow G12 8RZ, UK

{jana, jj}@dcs.gla.ac.uk

Abstract. We have proposed a novel image retrieval system that incorporates a workspace where users can organise their search results. A task-oriented and user-centred experiment has been devised involving design professionals and several types of realistic search tasks. We study the workspace's effect on two aspects: task conceptualisation and query formulation. A traditional relevance feedback system serves as baseline. The results of this study show that the workspace is more useful with respect to both of the above aspects. The proposed approach leads to a more effective and enjoyable search experience.

1 Introduction and Motivation

Content-based image retrieval (CBIR) systems have still not managed to find favour with the public even after more than a decade of research effort in the field. There are two main reasons for their lack of acceptability: first, the low-level features used to represent images in the system do not reflect the high-level concepts the user has in mind when looking at an image (*semantic gap*); and—partially due to this—the user tends to have major difficulties in formulating and communicating their information need effectively (*query formulation problem*).

We are seeking to find a solution to these problems by supporting an alternative search strategy. We have designed a system, *EGO*, that combines the search and the management process [1]. This is accomplished by introducing a workspace and recommendation system. While searching for images, the creation of groupings of related images is supported, encouraging the user to break up the task into related facets to organise their ideas and concepts. The system can then assist the user by recommending relevant images for selected groups. This way, the user can concentrate on solving specific tasks rather than having to think about how to create a good query in accordance with the retrieval mechanism.

Although a workspace has been introduced in a few IR systems before, e.g. Sketch-Trieve [2], its usefulness—especially for *image retrieval*—has not been evaluated formally yet. To remedy this shortcoming, we have designed a user experiment to evaluate the effectiveness of our approach for solving realistic image search tasks. We compare *EGO*'s performance to that of a traditional relevance feedback system as a baseline. In the relevance feedback system, the user is given the option of selecting relevant images from the search results in order to improve the results in the next iteration. Our aim is

to collect evidence on the systems' effectiveness as perceived by the users. More importantly, however, we would like to determine the workspace's role in helping the user to both conceptualise their search tasks and overcome the query formulation problem.

The experiment has been completed in two stages. Experiment 1 involved 12 participants using the two systems for category search tasks and a design task. A summary of the results has been published elsewhere [3]. This previous study only provided indicative conclusions on our research hypotheses. The results of the system's effectiveness were ambiguous: people performed better (using quantitative measures) in the category search tasks on the baseline while they were generally more satisfied with the workspace system. On the other hand, it indicated that the organisation did indeed help to conceptualise the tasks. However, we only studied two kinds of tasks from which the design task was only performed on the workspace system. So to be able to further study the effect of task on searching and organisation behaviour we need to investigate a larger variety of tasks. Above all, we failed to capture the grouping process' effect on the query formulation process, which we address in this paper.

In addition, the retrieval mechanism based on visual features only was limiting: the results were generally poor, the recommendation system was even more affected as it only returned the top 10 results, and the users were often swamped by irrelevant images due to not being able to provide negative feedback. We have decided to remedy these problems and introduce a different set of tasks for the second stage of the experiment. With this improved evaluation setup, Experiment 2 should help to clarify the validity of our research hypotheses. These results are presented in this paper. We look at the results from slightly different angles, such as a detailed analysis of the users' perception of task performance and user effort to complete the tasks. In particular, a direct comparison between the relevance feedback process and the grouping and recommendation system finally enables us to compare their impact on query formulation.

2 The Interfaces

We describe the main characteristics of both interfaces used in the evaluation, followed by an overview of the underlying retrieval system in this section.

2.1 Workspace Interface – WS

The interface used in the evaluation is a simplified version of that of the *EGO* system [1]. *EGO* has some additional features for personalisation and can, in principle, accommodate any sort of query facility. Since our main objective in these experiments is to evaluate the usefulness of the workspace, this interface is referred to as the Workspace Interface (*WS*). The *WS* interface depicted in Figure 1 comprises the following components:

1. Query Panel: This provides a basic query facility to search the database by allowing the user to compose a search request by entering search terms or adding example images to the query-by-example (QBE) panel provided here. Clicking on the "Search" button in this panel will issue a search.

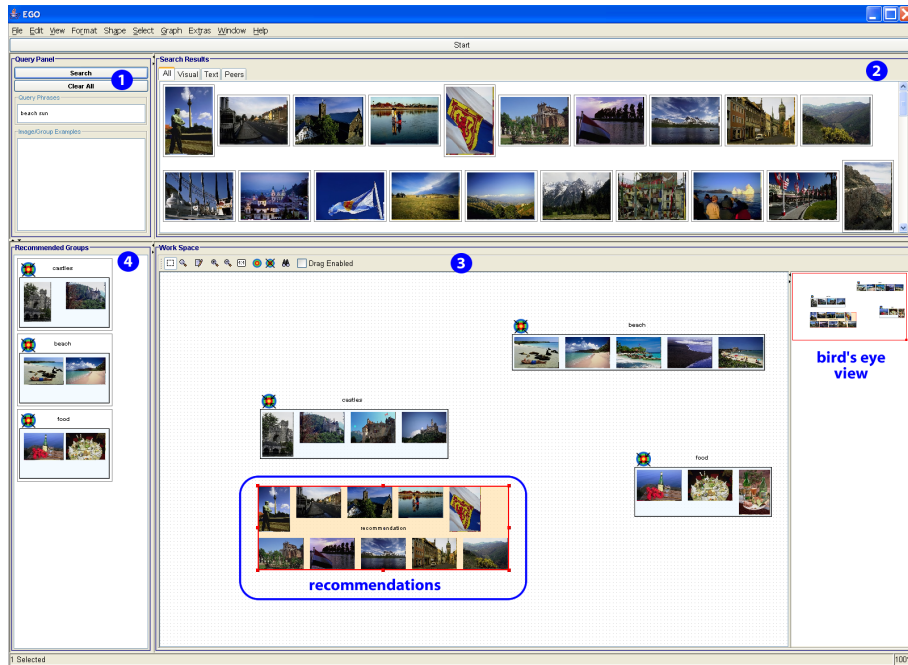


Fig. 1. Annotated WS interface

2. Results Panel: The search results from a query constructed in the Query Panel will be displayed in this panel. Any of the returned images can be dragged onto the workspace to start organising the collection or into the QBE panel to change the current query.
3. Workspace Panel: The workspace holds all the images added to it by the user, and serves as an organisation ground for the user to construct groupings of images. Groupings can be created by right-clicking anywhere on the workspace, which opens a context menu in which the option can be selected. Traditional drag-and-drop techniques allow the user to drag images into (or out of) a group or reposition the group on the workspace. An image can belong to multiple groups simultaneously. Panning and zooming techniques are supported to assist navigation in a large information space. The top 10 recommendations will be displayed close to the selected group on the workspace (see centre of workspace in Figure 1). In addition, the complete recommendation results are also displayed in the Results Panel.
4. Group Results Panel: For each query or recommendations issued the existing groups will be ranked in order of similarity to the current query/group and the five top matching groups will be displayed in this panel. Each returned group contains a link to the original group on the workspace.

To recapitulate, the query facilities available in the WS interface are: (1) manually constructed queries by providing keywords and/or one or more image examples (QBE), and (2) user-requested recommendations.

2.2 Relevance Feedback Interface – CS

The baseline system is a traditional relevance feedback system, referred to as *CS* (for Checkbox System). The interface contains the following components:

1. Query Panel: as above.
2. Results Panel: As above, but instead of dragging a relevant image onto the workspace the user has the choice of marking it as one of relevant (+), irrelevant (-), or neutral (=) by checking the respective combo box underneath the image. After images have been marked the user can ask the system to update the current search results (based on the feedback provided) by clicking the “Update Results” button in this panel.
3. Selected Items Panel: Any item selected relevant during the course of the search session will be added to this panel. The user can manually delete images from this panel if they change their mind at a later change.



Fig. 2. Relevance feedback in CS (image is marked “neutral”)

To summarise, the look-and-feel of the interface is similar to *WS* (without the workspace facility). Finally, *CS* supports two query facilities: (1) manual queries as above, and (2) automatic query reformulation by the feedback provided in the search results.

2.3 Retrieval System

The underlying retrieval system is the same in both interfaces and is described in [1]. Images are represented by a set of low-level visual features and modelled according to the hierarchical object model [4]. The distance between an object in the database and a given query representation is computed in two steps: computing the individual feature distances by the generalised Euclidean distance; then combining the individual distances linearly with a set of feature weights. The relevance feedback algorithm is implemented by an optimised framework for updating the retrieval parameters as proposed in [4]. It attempts to learn the best query representation and feature weighting for a selected group of images (positive training samples).

After the results obtained from the first set of participants, the experimental systems were scrutinised and consequently redesigned to take into account the lessons learnt. The main changes made were:

- The recommendation system in *WS* was not used to its full potential, due to its inability to recommend relevant images. This has been addressed in two ways. First, instead of just showing the top 10 recommendations on the workspace, the results panel now also shows the complete results (limited to 100 images). Second, a textual search facility has been introduced, because the visual features seemed not sufficient to solve more abstract tasks providing a more realistic search experience. Textual annotations obtained from [5] were incorporated and implemented according to the vector-space model [6]. Visual and textual features are combined using a rank-based list aggregation method [7].

- The retrieval mechanism was further improved by allowing negative feedback, as people complained about the inability to continue a search when the majority of returned images were irrelevant. Since incorporating negative feedback is a difficult endeavour [8], we have opted for a quick and safe approach: irrelevant images are added to a negative filter excluding them to be returned for the same search. It was straight-forward to implement this in CS where negative feedback can easily be provided explicitly. In WS however, we have chosen an implicit feedback strategy, whereby an image is automatically added to a negative filter for a group when it has been ignored (i.e. not dragged into this group) after having been returned 3 times amongst the top 10 recommendations.

3 Evaluation Methodology

It has been argued that traditional IR evaluation techniques based on precision-recall measures are not suitable for evaluating adaptive systems [9, 10]. Thus, we adopted a task-oriented, user-centred approach [10]. We have designed the experiments to be as close to real-life usage as possible: we have chosen participants with a design-related background and have set tasks that are practical and relevant.

We employed a subset of the Corel collection (CD 1, CD 4, CD 5, and CD 6 of the Corel 1.6M dataset), containing 12800 photographs in total. 12 searchers used two systems in a randomised within-subjects design. The independent variable was system type; two sets of values of a variety of dependent variables indicative of acceptability or user satisfaction were to be determined through questionnaires. In addition, users' actions were logged and analysed.

Participants. Since we wanted to test the system in a realistic usage scenario, our sample user population consisted of post-graduate design students and young design professionals. There were 12 participants, 7 male and 5 female, with a wide variety of ages in the range of 20-50 years. The average age was 28 years. The participants had on average 5 years experience in a design-related field (graphic design, architecture, photography). Most people dealt with digital images at least once a day as part of their course or work. The user profile in the current experiment is similar to Experiment 1.

Tasks. The tasks in Experiment 1 consisted of: category search tasks where participants had to find as many images as possible for a specific topic such as “elephants” (Task A) and a complex topic such as “underwater world” (Task B); and a design-task in which participants chose 3-5 images for designing a leaflet with a very open topic (Task C). We felt that more tasks were needed in order to draw definitive conclusions on the workspace's usefulness in helping to conceptualise tasks. We have devised a variety of realistic tasks, with different levels of complexity, abstraction and creativity. The tasks in Experiment 2 are:

Theme search task (Task D). In this task people were asked to find an image fitting into a specified theme (“people in national costumes” and “seasons in the country”). The theme was illustrated by three example images and the task involved searching for and selecting *one further image* complementing this set.

Illustration task (Task E). The task was to illustrate a piece of text for publication on the WWW, or an advertising slogan with *three images*. There were four tasks in total from which the participants had to choose two (one on each system).

Abstract search task (Task F). Here people were asked to select *at least one image* representing a given abstract topic (“cute” and “dynamic”). The simulated work task situation prescribed to select an image for a photo competition.

The level of abstraction grows from Task D to F: Task D has a very specified theme as defined by the three example images; the theme in Task E is specified by the given text or slogan but leaves more room for choosing the appropriate images; Task F finally leaves both the interpretation of the topic as well as the choice of images illustrating that topic open to the individual. Tasks A and B would be found before Task D on this scale, while Task C is a little more abstract and creative than Task E (since there was no text for illustration purposes). The complexity of a task is influenced by the coherence of its topic: if a topic is composed of several concepts or interpretations we regard it as complex. When we talk about facets of a task, we mean the concepts or interpretations that an individual filtered out when pursuing the task.

In Tasks D and F, the participants were assigned a specific topic, which was rotated based on a Latin-square design. The 4 topics of Task E were very similar in nature (level of complexity and available choice of suitable images). We have not found any significant variations between these topical alternatives within this search task group.

No time limits were set on the new tasks, as was learnt from Experiment 1 that this adversely affected people’s performance.

Hypothesis. Evidence was to be collected for the following sub-hypotheses:

1. The workspace leads to an increased effectiveness and user satisfaction.
2. The workspace helps to conceptualise and diversify tasks.
3. The grouping process helps to overcome the query formulation problem.

Procedure. Each participant performed four search sessions, using each system twice, completing two tasks with a different topic per system. The procedure involved a pre-search questionnaire, the four search sessions followed by a post-search questionnaire each time, and finally an exit questionnaire/interview comparing the systems. A search session was preceded by a training session if the system was used for the first time. The whole procedure lasted approximately two hours. Tasks and systems were rotated according to a Latin-square design in order to compensate the learning bias.

4 Results Analysis

The systems are compared according to (a) their effectiveness, and (b) user satisfaction. The following results are based on 48 (12×4) searches in total.

The results for Likert-scales and semantic differentials are in the range [1,5], the higher the value the better. Statistically significant differences are provided where appropriate with $p \leq .05$ using the two-tailed version of the non-parametric Wilcoxon Paired-Sample test. \overline{CS} and \overline{WS} denote the means for CS and WS respectively, while \widetilde{CS} and \widetilde{WS} denote the medians.

4.1 Effectiveness

The systems’ effectiveness is investigated from two sides: objectively from the perspective of the required effort as determined from the usage logs and subjectively from the perspective of the participants.

User Effort. Due to the lack of ground truth for the tasks in Experiment 2, we provide an analysis of the number of images selected per task and the amount of user effort required to select them. These include: total search time and number of queries issued. People can issue either manual queries—constructed in textual form, by providing image examples or a combination of both—or relevance feedback queries. The latter correspond to relevance feedback iterations in CS or group recommendations in WS.

Table 1 reveals that less queries were issued and more images were selected on WS. In particular, more *relevance feedback* queries were requested on WS, while more *manual queries* were constructed on CS (with the exception of Task F). The RF queries were particularly useful for Tasks E and F. The search session lasted on average longer on WS. By contrast, Task E stands out for being completed in less time on WS (with a difference of about 4 minutes) but still achieving a slightly larger selection of images in the end. This indicates that WS is particularly useful for design-oriented tasks.

Table 1. User effort indicators per task and system in Experiment 2

	<i>D_{CS}</i>	<i>D_{WS}</i>	<i>E_{CS}</i>	<i>E_{WS}</i>	<i>F_{CS}</i>	<i>F_{WS}</i>	<i>CS</i>	<i>WS</i>
time	9’55’’	12’02’’	18’26’’	14’18’’	9’40’’	14’31’’	12’40’’	13’35’’
#images	9.6	12.3	17.9	18.6	13.6	17.8	13.7	16.2
#queries	11.9	9.8	21.5	19.1	15.9	17.1	16.4	15.3
manual	8.4	7.6	15.8	12.2	11.4	12.0	11.9	10.6
RF	3.4	2.1	5.8	6.9	4.5	5.1	4.6	4.7

User Perception of Performance. After each task the users were asked if they thought they had succeeded in their performance of the task and also rate potential problems that might have affected their performance. Performing a task on WS was more successful, as can be seen in Table 2. In comparison to CS, people had a slight difficulty in

Table 2. User perception of task performance per task and system (performance:higher=better, problems:lower=more problematic)

	<i>D_{CS}</i>	<i>D_{WS}</i>	<i>E_{CS}</i>	<i>E_{WS}</i>	<i>F_{CS}</i>	<i>F_{WS}</i>	<i>CS</i>	<i>WS</i>	p
performance success	4.2	4.6	4.1	4.2	4.3	4.4	4.2	4.4	-
did not understand task	5.0	4.8	5.0	4.9	4.9	4.6	5.0	4.8	-
images not in collection	4.2	4.4	3.5	3.5	4.1	3.4	3.9	3.8	-
no relevant images returned	4.0	4.4	3.6	3.5	4.5	4.4	4.0	4.1	-
not enough time	4.8	4.9	4.4	4.1	4.9	4.6	4.7	4.5	-
unsure of next action	4.1	4.5	4.4	4.4	4.0	4.3	4.2	4.4	-

understanding the search task. Also, time was more of an issue on WS than CS¹. This reflects the increased cognitive effort required to perform a task on WS.

On the other hand, people's performance was hindered more by an uncertainty of what action to take next on CS. Together with the user comments presented below this indicates that—though a simple concept in principle—providing relevance feedback brings uncertainty as to which images to select for feedback in order to achieve better results. This corroborates similar results in textual information retrieval [11].

4.2 User Satisfaction

In this section we first present the user satisfaction with the particular interface features—the grouping and recommendation system in WS and the relevance feedback process in CS. After that, we discuss the responses concerning user satisfaction with the system in general.

Interface Support. People were asked how effective they found the interface and rated the contributing features. Table 3 summarises these results. Overall, WS was regarded significantly more *effective*. The three top rated features on WS were that it helped to *organise images*, *explore the collection*, and *analyse the task*. The ordering of features on CS was: *find relevant images*, *explore the collection*, and *detect/express different task aspects*. Apart from *find relevant images*, all features are rated significantly higher on WS.

Table 4 compares the adaptive querying mechanisms: the relevance assessment in CS and the grouping in WS. It turns out that the grouping was considered significantly more *effective* and *useful*. It is also interesting to note that the relevance assessment was even considered more *difficult* than the grouping.

In open-ended questions the participants were asked to state the most and least useful tools of the interface. The most useful tools in CS were stated as, in order of frequency of responses: textual query (10 responses²), QBE (9), and relevance feedback facility (7). The least useful tools were: result filters for various features (5), relevance feedback (4), and lack of storing facility/overview of selected images (4). Users who thought the relevance feedback was a useful tool stated it as helping them to improve and/or narrow down their search. The problems with relevance assessment were mainly that it returned unexpected results and that it was difficult to keep track of what the system was doing.

In WS, people unanimously liked the grouping facility on the workspace. The three most useful tools in WS included the grouping of images (14), group recommendations (10) and textual queries (5), and the least useful tools were: QBE (4), top 10 window of recommendations (3) and text search (2). This shows that using groups and recommendations was considered more useful than the manual search facilities; especially the query-by-example facility was superfluous in this system. The grouping's only disadvantage that became apparent was that it was difficult to remove images from existing groups.

These results support our view that WS, with its grouping and recommendation facility, assists the user in the query formulation process, while removing the need to

¹ In Experiment 1, people also tended to agree more with the statement that they had enough time to complete their task in CS: $\overline{CS} = 4.6$, and $\overline{WS} = 4.3$.

² This question was asked after each search, thus 24 responses are possible per system.

Table 3. Interface effectiveness

Statement	CS	CS	WS	WS	p
effective	3.7	4	4.4	5	0.032
analyse task	2.8	3	4.3	5	0.001
explore collection	3.5	4	4.6	5	0.001
find relevant images	4.2	4	4.2	4	-
organise images	2.7	3	4.7	5	0.001
detect/express task aspects	3.0	3	4.2	4	0.003

Table 4. Relevance assessment on CS vs. grouping on WS

Differential	CS	WS	p
easy	3.8	4.4	-
effective	3.3	4.3	0.019
useful	3.7	4.4	0.017

Table 5. Results for system part

Differential	CS	CS	WS	WS	p
wonderful	3.3	3	4.1	4	-
satisfying	3.2	3	4.0	4	-
stimulating	3.5	3	4.3	4	-
easy	4.0	4	3.8	4	-
flexible	2.9	3	4.2	4	0.004
efficient	3.3	3	3.9	4	-
novel	3.7	4	4.4	5	-

	CS	CS	WS	WS	p
in control	3.6	4	3.6	4	-
comfortable	3.7	4	4.3	5	-
confident	3.1	3	3.8	4	-
learn to use	4.1	4	3.9	4	-
use	3.9	4	3.9	4	-

manually reformulate queries. The picture in CS is quite different: people were divided on the usefulness of the relevance assessments and some still relied heavily on the manual query facilities. On average, people selected 2.4, 3.2, and 3.8 images per relevance feedback iteration for Task D, E, and F, respectively. Compared to that, the groups in WS contained 4.9, 4.6, and 4.4 images. So the manual selection process was less productive than collecting the images in groups. Moreover, the grouping process has the additional benefit of supporting a diversifying search by allowing to declare and pursue various task aspects simultaneously.

System. In the post-search questionnaires, the participants considered CS more *easy* than WS, while they considered WS to be significantly more *flexible*. The scores for the remaining differentials, *wonderful*, *satisfying*, *stimulating*, *efficient*, and *novel* were generally higher for WS as well (see Table 5). While using the system, people felt more *comfortable* and *confident*. However, WS was more difficult to *learn to use*.

After completing all four search tasks, the users were asked to determine the system that was (a) easiest to learn to use, (b) easiest to use, (c) most effective, and (d) they liked best overall in the exit questionnaire. They could choose between WS, CS, and no difference as responses to these questions. 67% liked WS best and the majority also thought it was more effective (46% compared to 26% for CS). CS was clearly easier to learn to use (58%), whereas the ranking for using the systems was relatively balanced (46% for WS and 42% for CS).

Finally, the participants were asked for their opinion on what they liked or disliked about each system. The responses reconfirmed most advantages and disadvantages already identified in the previous experiment. The advantages listed for CS were that it was easy to use, fast and efficient especially for specific searches. Its disadvantages

included that the users felt they did not have enough control over the search and that its interface and search process was less intuitive.

People appreciated WS as an organising tool. The workspace enabled them to plan their tasks and pursue alternative search threads, without losing the overview of intermediate results and searches. Once more, the system was regarded as more flexible and offering better control over the search process. In Experiment 1, the disadvantages were mainly concerned with the poor quality of the recommendations and that the handling of groups was sometimes cumbersome. Both of these issues are not inherent in the interaction paradigm of the proposed system itself, and were consequently improved for Experiment 2. The recommendation quality was improved by taking textual annotations into account. The handling of the groups and images within groups was changed so that the system now automatically arranges the layout of the images in a group. Consequently, none of these issues resurfaced in Experiment 2.

Organisation Analysis. A further objective of this study was to see if there is any correlation between task characteristics and the way people organise images on the workspace. Due to space restrictions we can only briefly provide our conclusions here. We observed that the more open or complex a task is, the more groups were created on the workspace (1.5, 2.9, and 2.6 for Task D, E, and F, respectively). For these types of tasks the organisation was deemed most useful and recommendations were requested more often.

The groups the participants created for any given task often overlapped in the overall themes of the groups, but not necessarily the images themselves. This shows that groups are to a great extent task-dependent and hence people would possibly benefit from using and working with other people's groups. We briefly list some examples of the facets created by the participants for two topics. For Task D, topic "seasons in the country", people created groups for "autumn" images (the image that was missing from the set) mainly displaying leafy, red forests; other groups created were "colourful fields", "close-up of plants", "boats", "country houses". The abstract topic "dynamic" of Task E was illustrated by the following groups: "animals" (sometimes split into "flying birds", "tigers/leopards", etc.), "sports", "mountains", "waterfalls", "sunsets/landscapes", "boats/water".

5 Discussion

In this section, we explore the benefits of the workspace system in comparison to the traditional relevance feedback approach. Our observations are based on the overall results of the two-stage experiment involving 24 participants on a variety of realistic search tasks.

First, we investigated the systems' ability to support the users in solving their tasks. The questionnaire responses indicated that the workspace helped them to analyse and explore their tasks better. This is most likely as a result of it allowing them to explore the facets of the search task they were performing. Together with the recommendation facility, this has increased the effectiveness of the system. The required effort to complete a task was lower on WS: less queries were issued to find a larger selection of images.

In particular, users created less manual queries but issued more system recommendations. The participants also perceived their performance as more successful on WS and the interface was perceived significantly more effective for completing the tasks. This shows that the workspace helped to *conceptualise and diversify the task* better and as a result *increased the effectiveness of the search*.

Moreover, the grouping facility was not only considered easier, more effective and useful than the relevance feedback approach in CS, but was praised unanimously in open-ended questions. In addition, the relevance feedback facility caused more confusion. It became apparent that providing relevance feedback brings uncertainty as to which images to select for feedback in order to improve the results. Hence, people relied more on the manual query facilities on CS than WS. Although both systems have the same underlying retrieval mechanism, the workspace approach is more successful at eliciting constructive feedback while hiding the internals of the retrieval mechanism. It is more natural to the user to provide feedback in a structured form by creating groups on the workspace instead of indicating relevant and irrelevant images indiscriminately. The groups allowed users to see which images contributed to the result list. Consequently, people selected more images for feedback and requested more recommendations on WS than RF iterations on CS. Thus, one can conclude that the grouping process is *better at overcoming the query formulation problem*.

We also found a link between task and the use of the workspace. The more complex or open the task, the more useful the workspace was perceived to be. For these tasks, the organisation was regarded more useful and recommendations were consulted more often. With growing task complexity, users created more groups which allowed them to explore the task and collection by following up on various facets (trains of thought). On the other hand, CS was better for tasks that required selection of a large number of images for a very specific topic.

These observations have led us to accept all three experimental hypotheses. However, this study also helped to identify the limitations of the workspace. WS was more difficult to use and the cognitive effort required to solve a task was higher. This was reflected in the questionnaire responses; in particular users had more difficulty in understanding the task and it took longer to complete it. However, the longer learning period and increased cognitive effort is not perceived as a disadvantage of WS; after all, 16 people preferred WS over CS. More importantly, we found evidence that attributed the prolonged search session to the system's ability to support the user in exploring the tasks from different perspectives. As mentioned before, people were able to diversify their search better and follow up on multiple trains of thought simultaneously. Still, one has to keep in mind that it takes longer to become familiarised with this interface, although we strived to make its operation as intuitive as possible by using standard commands which the user may already be familiar with wherever possible.

Finally, we did not explore the use of WS for collaborative image retrieval. On the workspace, people leave footprints of their activities behind for later usage. We also observed that people's groups overlapped in their overall themes, which could be exploited in a collaborative context. Such a feature will be explored in future studies.

6 Conclusion

In this paper, we have established the usefulness of the workspace system for image retrieval. We have created a realistic experimental study, in which design professionals performed a variety of realistic search tasks. Based on the results of this experiment, we argue that the workspace is an indispensable tool in an image retrieval system. It is used for organising the results according to the different aspects or facets of the task. This helps users greatly in analysing and exploring the task as well as the collection. Moreover, the workspace supports a more intuitive search process and helps to overcome the query formulation problem. All these factors lead to a more effective and enjoyable search experience.

References

1. Urban, J., Jose, J.M.: EGO: A personalised multimedia management and retrieval tool. *International Journal of Intelligent Systems (IJIS)*, Special Issue on 'Intelligent Multimedia Retrieval' (2005) to appear.
2. Hendry, D.G., Harper, D.J.: An informal information-seeking environment. *Journal of the American Society for Information Science* **48** (1997) 1036–1048
3. Urban, J., Jose, J.M.: An explorative study of interface support for image searching. In: *Proc. of the 3rd Int. Workshop on Adaptive Multimedia Retrieval*. (2005)
4. Rui, Y., Huang, T.S.: Optimizing learning in image retrieval. In: *IEEE Proc. of Conf. on Computer Vision and Pattern Recognition (CVPR-00)*, IEEE Computer Society Press (2000) 236–245
5. Berkley's Digital Library Project: <http://elib.cs.berkeley.edu/photos/corel/> (2005)
6. Salton, G., McGill, M.J.: *Introduction to Modern Information Retrieval*. McGraw-Hill, Tokio (1983)
7. Urban, J., Jose, J.M.: Evidence combination for multi-point query learning in content-based image retrieval. In: *Proc. of the IEEE Sixth Int. Symposium on Multimedia Software Engineering (ISMSE'04)*. (2004) 583–586
8. Zhou, X.S., Huang, T.: Relevance feedback in image retrieval: A comprehensive review. *ACM Multimedia Systems Journal* **8** (2003) 536–544
9. Jose, J.M., Furner, J., Harper, D.J.: Spatial querying for image retrieval: A user-oriented evaluation. In: *Proc. of the Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR'98)*, ACM Press (1998) 232–240
10. Ingwersen, P.: *Information Retrieval Interaction*. Taylor Graham, London (1992)
11. Beaulieu, M., Jones, S.: Interactive searching and interface issues in the Okapi best match probabilistic retrieval system. *Interacting with Computers* **10** (1998) 237–248